RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: _10/551,057_

Source: _PUT/o_

Date Processed by STIC: _10/7/05_

ENTERED

PCT

RAW SEQUENCE LISTING   DATE: 10/07/2005
PATENT APPLICATION: **US/10/551,057**   TIME: 10:19:17

Input Set : **A:\21188P SEQLIST.TXT**
Output Set: **N:\CRF4\10072005\J551057.raw**

```
   4 <110> APPLICANT: Jansen, Kathrin U.
   5         Schultz, Loren D.
   6         Neeper, Michael P.
   7         Markus, Henry Z.
   9 <120> TITLE OF INVENTION: OPTIMIZED EXPRESSION OF HPV 31 L1 IN
  10         YEAST
  12 <130> FILE REFERENCE: 21188P
C--> 14 <140> CURRENT APPLICATION NUMBER: US/10/551,057
C--> 14 <141> CURRENT FILING DATE: 2005-09-26
  14 <150> PRIOR APPLICATION NUMBER: PCT/US2004/008677
  15 <151> PRIOR FILING DATE: 2004-03-19
  17 <150> PRIOR APPLICATION NUMBER: 60/457,172
  18 <151> PRIOR FILING DATE: 2003-03-23
  20 <160> NUMBER OF SEQ ID NOS: 8
  22 <170> SOFTWARE: FastSEQ for Windows Version 4.0
  24 <210> SEQ ID NO: 1
  25 <211> LENGTH: 1515
  26 <212> TYPE: DNA
  27 <213> ORGANISM: HPV31 L1 wild-type
  29 <400> SEQUENCE: 1
  30 atgtctctgt ggcggcctag cgaggctact gtctacttac cacctgtccc agtgtctaaa 60
  31 gttgtaagca cggatgaata tgtaacacga accaacatat attatcacgc aggcagtgct 120
  32 aggctgctta cagtaggcca tccatattat tccataccta aatctgacaa tcctaaaaaa 180
  33 atagttgtac caaaggtgtc aggattacaa tatagggtat ttagggttcg tttaccagat 240
  34 ccaaacaaat ttggatttcc tgatacatct ttttataatc ctgaaactca acgcttagtt 300
  35 tgggcctgtg ttggtttaga ggtaggtcgc gggcagccat taggtgtagg tattagtggt 360
  36 catccattat taaataaatt tgatgacact gaaaactcta atagatatgc cggtggtcct 420
  37 ggcactgata atagggaatg tatatcaatg gattataaac aaacacaact gtgtttactt 480
  38 ggttgcaaac cacctattgg agagcattgg ggtaaaggta gtccttgtag taacaatgct 540
  39 attaccccctg gtgattgtcc tccattagaa ttaaaaaatt cagttataca agatggggat 600
  40 atggttgata caggctttgg agctatggat tttactgctt tacaagacac taaaagtaat 660
  41 gttccttttgg acatttgtaa ttctatttgt aaatatccag attatcttaa aatggttgct 720
  42 gagccatatg gcgatacatt attttttatt ttacgtaggg aacaaatgtt tgtaaggcat 780
  43 ttttttaata gatcaggcac ggttggtgaa tcggtcccta ctgacttata tattaaaggc 840
  44 tccggttcaa cagctacttt agctaacagt acatactttc ctacacctag cggctccatg 900
  45 gttacttcag atgcacaaat ttttaataaa ccatattgga tgcaacgtgc tcagggacac 960
  46 aataatggta tttgttgggg caatcagtta tttgttactg tggtagatac cacacgtagt 1020
  47 accaatatgt ctgtttgtgc tgcaattgca aacagtgata ctacatttaa aagtagtaat 1080
  48 tttaaagagt atttaagaca tggtgaggaa tttgatttac aatttatatt tcagttatgc 1140
  49 aaaataacat tatctgcaga cataatgaca tatattcaca gtatgaatcc tgctatttttg 1200
  50 gaagattgga attttggatt gaccacacct ccctcaggtt ctttggagga tacctatagg 1260
  51 tttgtaacct cacaggccat tacatgtcaa aaaagtgccc cccaaaagcc caaggaagat 1320
  52 ccatttaaag attatgtatt ttgggaggtt aatttaaaag aaaagtttttc tgcagattta 1380
```

RAW SEQUENCE LISTING   DATE: 10/07/2005
PATENT APPLICATION: US/10/551,057   TIME: 10:19:17

Input Set : A:\21188P SEQLIST.TXT
Output Set: N:\CRF4\10072005\J551057.raw

```
53 gatcagtttc cactgggtcg caaatttta ttacaggcag gatatagggc acgtcctaaa 1440
54 tttaaagcag gtaaacgtag tgcaccctca gcatctacca ctacaccagc aaaacgtaaa 1500
55 aaaactaaaa agtaa                                                1515
57 <210> SEQ ID NO: 2
58 <211> LENGTH: 1515
59 <212> TYPE: DNA
60 <213> ORGANISM: Artificial Sequence
62 <220> FEATURE:
63 <223> OTHER INFORMATION: 31 partial rebuild
65 <400> SEQUENCE: 2
66 atgtctctgt ggcggcctag cgaggctact gtctacttac cacctgtccc agtgtctaaa 60
67 gttgtaagca cggatgaata tgtaacacga accaacatat attatcacgc aggcagtgct 120
68 aggctgctta cagtaggcca tccatattat tccataccta aatctgacaa tcctaaaaaa 180
69 atagttgtac caaaggtgtc aggattacaa tatagggtat ttagggttcg tttaccagat 240
70 ccaaacaaat ttggatttcc tgatacatct tttttataatc ctgaaactca acgcttagtt 300
71 tgggcctgtg ttggtttaga ggtaggtcgc gggcagccat taggtgtagg tattagtggt 360
72 catccattat taaataaatt tgatgacact gaaaactcta atagatatgc cggtggtcct 420
73 ggcactgata atagggaatg tatatcaatg gattataaac aaacacaact gtgtttactt 480
74 ggttgcaaac cacctattgg agagcattgg ggtaaaggta gtccttgtag taacaatgct 540
75 attacccctg gtgattgtcc tccattagaa ttaaaaaatt cagttataca agatggggat 600
76 atggttgata caggctttgg agctatggat tttactgctt tacaagacac taaaagtaat 660
77 gttcctttgg acatttgtaa ttctatttgt aaatatccag attatcttaa aatggttgct 720
78 gagccatacg gcgacacctt gttcttctat ttgcgtagag aacagatgtt cgtaaggcac 780
79 ttcttcaaca gatccggcac cgtaggtgaa tctgtcccaa ccgacctgta catcaagggc 840
80 tccggttcca ccgctaccct ggctaactcc acctacttcc caactccatc tggctccatg 900
81 gtcacctccg acgctcagat cttcaacaag ccatactgga tgcagcgtgc acaggtcac 960
82 aacaacggta tctgttgggg taaccagctg ttcgtgactg tggtcgatac cacgcgttct 1020
83 accaacatgt ctgtctgtgc tgcaatcgct aactctgaca ctaccttcaa gtcctctaac 1080
84 ttcaaggagt acctgagaca tggtgaggaa ttcgatctgc aattcatctt ccagttgtgc 1140
85 aagatcaccc tgtctgctga catcatgacc tacatccaca gtatgaaccc tgccatcctg 1200
86 gaggactgga acttcggtct gaccactcca ccttccggtt ctttggagga tacctatagg 1260
87 tttgtaacct cacaggccat tacatgtcaa aaaagtgccc cccaaaagcc caaggaagat 1320
88 ccatttaaag attatgtatt ttgggaggtt aatttaaaag aaaagttttc tgcagattta 1380
89 gatcagtttc cactgggtcg caaatttta ttacaggcag gatatagggc acgtcctaaa 1440
90 tttaaagcag gtaaacgtag tgcaccctca gcatctacca ctacaccagc aaaacgtaaa 1500
91 aaaactaaaa agtaa                                                1515
93 <210> SEQ ID NO: 3
94 <211> LENGTH: 1515
95 <212> TYPE: DNA
96 <213> ORGANISM: Artificial Sequence
98 <220> FEATURE:
99 <223> OTHER INFORMATION: 31 total rebuild
101 <400> SEQUENCE: 3
102 atgtctttgt ggagaccatc tgaagctacc gtctacttgc caccagtccc agtctctaag 60
103 gtcgtctcta ccgacgaata cgtcaccaga accaacatct actaccacgc tggttctgct 120
104 agattgttga ccgtcggtca cccatactac tctatcccaa agtctgacaa cccaaagaag 180
105 atcgtcgtcc caaaggtctc tggtttgcaa tacagagtct tcagagtcag attgccagac 240
106 ccaaacaagt tcggtttccc agacacctct ttctacaacc cagaaaccca agattggtc 300
107 tgggcttgtg tcggtttgga agtcggtaga ggtcaaccat ggggtgtcgg tatctctggt 360
```

```
108 cacccattgt tgaacaagtt cgacgacacc gaaaactcta acagatacgc tggtggtcca 420
109 ggtaccgaca acagagaatg tatctctatg gactacaagc aaacccaatt gtgtttgttg 480
110 ggttgtaagc caccaatcgg tgaacactgg ggtaagggtt ctccatgttc taacaacgct 540
111 atcacccag gtgactgtcc accattggaa ttgaagaact ctgtcatcca agacggtgac 600
112 atggtcgaca ccggtttcgg tgctatggac ttcaccgctt tgcaagacac caagtctaac 660
113 gtcccattgg acatctgtaa ctctatctgt aagtacccag actacttgaa gatggtcgct 720
114 gaaccatacg gcgacacctt gttcttctac ttgcgtagag aacagatgtt cgtaaggcac 780
115 ttcttcaaca gatccggcac cgtaggtgaa tctgtcccaa ccgacctgta catcaagggc 840
116 tccggttcca ccgctaccct ggctaactcc acctacttcc caactccatc tggctccatg 900
117 gtcacctccg acgctcagat cttcaacaag ccatactgga tgcagcgtgc acagggtcac 960
118 aacaacggta tctgttgggg taaccagctg ttcgtgactg tggtcgatac cacgcgttct 1020
119 accaacatgt ctgtctgtgc tgcaatcgct aactctgaca ctaccttcaa gtcctctaac 1080
120 ttcaaggagt acctgagaca tggtgaggaa ttcgatctgc aattcatctt ccagttgtgc 1140
121 aagatcaccc tgtctgctga catcatgacc tacatccaca gtatgaaccc tgccatcctg 1200
122 gaggactgga acttcggtct gaccactcca ccttccggtt ctttggaaga cacctacaga 1260
123 ttcgtcacct ctcaagctat cacctgtcaa aagtctgctc cacaaaagcc aaaggaagac 1320
124 ccattcaagg actacgtctt ctgggaagtc aacttgaagg aaaagttctc tgctgacttg 1380
125 gaccaattcc cattgggtag aaagttcttg ttgcaagctg gttacagagc tagaccaaag 1440
126 ttcaaggctg gtaagagatc tgctccatct gcttctacca ccaccccagc taagagaaag 1500
127 aagaccaaga agtaa 1515
129 <210> SEQ ID NO: 4
130 <211> LENGTH: 504
131 <212> TYPE: PRT
132 <213> ORGANISM: Artificial Sequence
134 <220> FEATURE:
135 <223> OTHER INFORMATION: HPV 31 L1
137 <400> SEQUENCE: 4
138 Met Ser Leu Trp Arg Pro Ser Glu Ala Thr Val Tyr Leu Pro Pro Val
139 1               5                  10                  15
140 Pro Val Ser Lys Val Val Ser Thr Asp Glu Tyr Val Thr Arg Thr Asn
141            20                  25                  30
142 Ile Tyr Tyr His Ala Gly Ser Ala Arg Leu Leu Thr Val Gly His Pro
143        35                  40                  45
144 Tyr Tyr Ser Ile Pro Lys Ser Asp Asn Pro Lys Lys Ile Val Val Pro
145    50                  55                  60
146 Lys Val Ser Gly Leu Gln Tyr Arg Val Phe Arg Val Arg Leu Pro Asp
147 65                  70                  75                  80
148 Pro Asn Lys Phe Gly Phe Pro Asp Thr Ser Phe Tyr Asn Pro Glu Thr
149                85                  90                  95
150 Gln Arg Leu Val Trp Ala Cys Val Gly Leu Glu Val Gly Arg Gly Gln
151            100                 105                 110
152 Pro Leu Gly Val Gly Ile Ser Gly His Pro Leu Leu Asn Lys Phe Asp
153        115                 120                 125
154 Asp Thr Glu Asn Ser Asn Arg Tyr Ala Gly Gly Pro Gly Thr Asp Asn
155    130                 135                 140
156 Arg Glu Cys Ile Ser Met Asp Tyr Lys Gln Thr Gln Leu Cys Leu Leu
157 145                 150                 155                 160
158 Gly Cys Lys Pro Pro Ile Gly Glu His Trp Gly Lys Gly Ser Pro Cys
159                165                 170                 175
```

RAW SEQUENCE LISTING   DATE: 10/07/2005
PATENT APPLICATION: US/10/551,057   TIME: 10:19:17

Input Set : A:\21188P SEQLIST.TXT
Output Set: N:\CRF4\10072005\J551057.raw

```
160 Ser Asn Asn Ala Ile Thr Pro Gly Asp Cys Pro Pro Leu Glu Leu Lys
161             180             185             190
162 Asn Ser Val Ile Gln Asp Gly Asp Met Val Asp Thr Gly Phe Gly Ala
163         195             200             205
164 Met Asp Phe Thr Ala Leu Gln Asp Thr Lys Ser Asn Val Pro Leu Asp
165     210             215             220
166 Ile Cys Asn Ser Ile Cys Lys Tyr Pro Asp Tyr Leu Lys Met Val Ala
167 225             230             235             240
168 Glu Pro Tyr Gly Asp Thr Leu Phe Phe Tyr Leu Arg Arg Glu Gln Met
169             245             250             255
170 Phe Val Arg His Phe Phe Asn Arg Ser Gly Thr Val Gly Glu Ser Val
171             260             265             270
172 Pro Thr Asp Leu Tyr Ile Lys Gly Ser Gly Ser Thr Ala Thr Leu Ala
173         275             280             285
174 Asn Ser Thr Tyr Phe Pro Thr Pro Ser Gly Ser Met Val Thr Ser Asp
175     290             295             300
176 Ala Gln Ile Phe Asn Lys Pro Tyr Trp Met Gln Arg Ala Gln Gly His
177 305             310             315             320
178 Asn Asn Gly Ile Cys Trp Gly Asn Gln Leu Phe Val Thr Val Val Asp
179             325             330             335
180 Thr Thr Arg Ser Thr Asn Met Ser Val Cys Ala Ala Ile Ala Asn Ser
181         340             345             350
182 Asp Thr Thr Phe Lys Ser Ser Asn Phe Lys Glu Tyr Leu Arg His Gly
183         355             360             365
184 Glu Glu Phe Asp Leu Gln Phe Ile Phe Gln Leu Cys Lys Ile Thr Leu
185     370             375             380
186 Ser Ala Asp Ile Met Thr Tyr Ile His Ser Met Asn Pro Ala Ile Leu
187 385             390             395             400
188 Glu Asp Trp Asn Phe Gly Leu Thr Thr Pro Pro Ser Gly Ser Leu Glu
189             405             410             415
190 Asp Thr Tyr Arg Phe Val Thr Ser Gln Ala Ile Thr Cys Gln Lys Ser
191             420             425             430
192 Ala Pro Gln Lys Pro Lys Glu Asp Pro Phe Lys Asp Tyr Val Phe Trp
193         435             440             445
194 Glu Val Asn Leu Lys Glu Lys Phe Ser Ala Asp Leu Asp Gln Phe Pro
195     450             455             460
196 Leu Gly Arg Lys Phe Leu Leu Gln Ala Gly Tyr Arg Ala Arg Pro Lys
197 465             470             475             480
198 Phe Lys Ala Gly Lys Arg Ser Ala Pro Ser Ala Ser Thr Thr Thr Pro
199             485             490             495
200 Ala Lys Arg Lys Lys Thr Lys Lys
201             500
204 <210> SEQ ID NO: 5
205 <211> LENGTH: 34
206 <212> TYPE: DNA
207 <213> ORGANISM: Artificial Sequence
209 <220> FEATURE:
210 <223> OTHER INFORMATION: PCR Primer
212 <400> SEQUENCE: 5
```

RAW SEQUENCE LISTING   DATE: 10/07/2005
PATENT APPLICATION: **US/10/551,057**   TIME: 10:19:17

Input Set : **A:\21188P SEQLIST.TXT**
Output Set: **N:\CRF4\10072005\J551057.raw**

```
213 cgtcgacgta aacgtgtatc atattttttt acag                    34
215 <210> SEQ ID NO: 6
216 <211> LENGTH: 25
217 <212> TYPE: DNA
218 <213> ORGANISM: Artificial Sequence
220 <220> FEATURE:
221 <223> OTHER INFORMATION: PCR Primer
223 <400> SEQUENCE: 6
224 cagacacatg tattacatac acaac                              25
226 <210> SEQ ID NO: 7
227 <211> LENGTH: 41
228 <212> TYPE: DNA
229 <213> ORGANISM: Artificial Sequence
231 <220> FEATURE:
232 <223> OTHER INFORMATION: PCR Primer
234 <400> SEQUENCE: 7
235 ctcagatctc acaaaacaaa atgtctctgt ggcggcctag c            41
237 <210> SEQ ID NO: 8
238 <211> LENGTH: 38
239 <212> TYPE: DNA
240 <213> ORGANISM: Artificial Sequence
242 <220> FEATURE:
243 <223> OTHER INFORMATION: PCR Primer
245 <400> SEQUENCE: 8
246 gacagatctt actttttagt ttttttacgt tttgctgg               38
```

VERIFICATION SUMMARY   DATE: 10/07/2005
PATENT APPLICATION: **US/10/551,057**   TIME: 10:19:18

Input Set : **A:\21188P SEQLIST.TXT**
Output Set: **N:\CRF4\10072005\J551057.raw**

L:14 M:270 C: Current Application Number differs, Replaced Current Application No
L:14 M:271 C: Current Filing Date differs, Replaced Current Filing Date